



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning-Based Prediction of Sleep Quality Using Smartphone Usage Patterns

Mayuri Mutkekar, Diksha Marathe, Priya Patil, Vaishnavi Mali, Pooja Kurne, Poonam Fegade

K.K. Wagh Institute of Engineering Education and Research

Master of Computer Applications (MCA) Department, Nashik, Maharashtra, India

Abstract—In today's world, many students and young adults use their smartphones extensively, raising concerns about the impact on sleep quality. Exposure to smartphone screens at night suppresses melatonin production due to blue light emission and increases cognitive stimulation, disrupting natural circadian rhythms. This study proposes a machine learning-based framework to predict sleep quality based on smartphone usage patterns collected from students aged 18 to 25. The dataset includes features such as daily screen time, nighttime usage duration, notification frequency, social media usage, and sleep duration. Three machine learning models were evaluated: Random Forest Regression, Linear Regression, and Support Vector Machine (SVM). The Random Forest model achieved the highest predictive accuracy with an R^2 score of 0.82, demonstrating its ability to capture complex behavioral relationships. Results confirm a strong association between excessive smartphone use and degraded sleep quality, including reduced sleep duration and delayed sleep onset. This research contributes to digital wellness by providing a data-driven system to monitor and improve sleep habits.

Keywords—Machine Learning, Sleep Quality Prediction, Smartphone Usage Patterns, Random Forest Regression, Predictive Modeling, Digital Wellness, Behavioral Analytics

I. INTRODUCTION

Smartphones have become an integral component of modern life, particularly among students and young adults, serving as essential tools for communication, education, and entertainment. However, the rapid increase in smartphone usage has raised significant concerns regarding its impact on human health, especially sleep quality. Recent studies indicate that excessive use of electronic devices is strongly associated with sleep disturbances and reduced overall well-being [1].

Sleep is a fundamental biological process necessary for maintaining cognitive performance, emotional stability, and physical health. Poor sleep quality can lead to fatigue, reduced concentration, impaired academic performance, and long-term health issues such as stress and anxiety. Among various contributing factors, excessive smartphone usage—especially during nighttime—has emerged as a major cause of sleep disruption.

One of the primary mechanisms behind this effect is the exposure to blue light emitted from smartphone screens, which suppresses melatonin production and disrupts the natural circadian rhythm, thereby delaying sleep onset [2]. In addition, prolonged engagement with social media platforms and frequent notifications increases cognitive and psychological stimulation, making it difficult for individuals to relax before sleep [3].

Although several studies have investigated the relationship between smartphone usage and sleep quality, most existing research focuses on descriptive or statistical analysis and lacks predictive modeling approaches. Furthermore, there is limited integration of machine learning techniques to analyze behavioral data and forecast sleep outcomes effectively.

To address these limitations, this study proposes a machine learning-based predictive framework that analyzes smartphone usage patterns to estimate sleep quality among students and young adults aged 18–25. The study employs multiple machine learning algorithms, including Random Forest Regression, Linear Regression, and Support Vector Machine (SVM), to identify key factors influencing sleep health and to develop an accurate prediction model.

The main contributions of this research are as follows:

- Development of a predictive model for sleep quality using smartphone usage data.
- Comparative analysis of multiple machine learning algorithms.
- Identification of key behavioral factors affecting sleep patterns.
- Contribution to digital wellness through data-driven insights.

II. PROBLEM STATEMENT

Excessive smartphone usage, particularly during nighttime, is a significant contributor to poor sleep quality among students and young adults. Prolonged screen exposure suppresses melatonin secretion and disrupts circadian rhythms, causing delayed sleep onset and reduced sleep duration [2]. Frequent notifications and sustained social media engagement maintain heightened cognitive arousal, making it difficult for individuals to transition into restful sleep [3].

Despite widespread recognition of this problem, most existing studies adopt a descriptive or correlational approach rather than predictive modeling. The use of machine learning techniques to analyze smartphone behavioral data and forecast sleep quality remains limited [1]. This study addresses this gap by developing a machine learning model capable of analyzing smartphone usage patterns and generating actionable insights about sleep health.

III. OBJECTIVES

The main goals of this study are as follows:

- To analyze smartphone usage behaviors of students and young adults aged 18–25, including daily screen time, nighttime usage frequency, and notification patterns.
- To examine the influence of smartphone usage on sleep outcomes, including sleep duration, sleep onset latency, and overall sleep quality.
- To develop a data-driven machine learning model that predicts sleep quality from smartphone usage features.
- To compare multiple machine learning algorithms—Random Forest Regression, Linear Regression, and Support Vector Machine—to identify the most effective approach.
- To determine which smartphone usage features most significantly impact sleep quality.
- To contribute to digital wellness by providing data-driven insights and encouraging healthier smartphone habits.

IV. LITERATURE REVIEW

Previous research has consistently demonstrated the adverse effects of electronic device usage on sleep, particularly among teenagers and young adults. Key findings from the literature are summarized below.

Cain and Gradisar (2010) established that extensive electronic media use significantly disrupts sleep patterns in adolescents, resulting in delayed bedtimes and reduced total sleep duration [1]. Chang et al. (2015) demonstrated that exposure to light-emitting devices in the evening suppresses melatonin synthesis, disrupts circadian timing, and impairs next-morning alertness [3].

Exelmans and Van den Bulck (2016) found a strong association between pre-sleep smartphone use and shorter sleep duration, increased sleep onset latency, and subjective sleep complaints [5]. Similarly, Levenson et al. (2017) reported that frequent social media engagement contributes to heightened cognitive arousal that interferes with the ability to relax before sleep [4].

Hale and Guan (2015) conducted a systematic review confirming that greater screen time is linked to sleep insufficiency across various age groups [2]. Thomée et al. (2011) further corroborated these findings, linking heavy mobile phone use to stress, sleep disturbances, and depressive symptoms in young adults [6].

Despite this substantial body of evidence, the majority of prior studies have relied on descriptive analysis or inferential statistics without employing predictive models. Machine learning has demonstrated significant utility in behavioral data analysis across numerous domains; however, its application to sleep quality prediction from smartphone usage data remains limited. This study addresses this gap by

developing and evaluating machine learning models for this purpose.

V. METHODOLOGY

A. Research Design and Data Collection

This study adopted a survey-based approach to collect behavioral data from students and young adults aged 18–25. Participants were asked to self-report their smartphone usage patterns and sleep outcomes through structured questionnaires. The dataset comprises the following features:

TABLE I: DESCRIPTION OF DATASET FEATURES

Feature	Description
Screen Time	Total smartphone usage per day (hours)
Nighttime Usage	Duration of smartphone use before sleep
Social Media Usage	Frequency of social media activity
Notification Frequency	Number of notifications received
Sleep Duration	Total sleep hours per night
Sleep Score Quality	User-rated sleep quality

B. Data Preprocessing

Raw data underwent the following preprocessing steps to ensure quality and consistency:

- Handling of missing values through imputation or removal.
- Normalization and standardization of numerical features.
- Encoding of categorical variables into numerical representations.
- Detection and removal of outliers to prevent model distortion.

C. Feature Selection

The following features were selected based on their established relevance to sleep quality: Screen Time, Nighttime Usage, Notification Frequency, Social Media Usage, and Sleep Duration. Feature selection was performed using correlation analysis and Recursive Feature Elimination (RFE) to identify the most predictive variables.

D. Machine Learning Models

Three machine learning models were evaluated in this study:

- Random Forest Regression: An ensemble method that constructs multiple decision trees and aggregates their outputs for robust and accurate prediction.
- Linear Regression: A baseline model that assumes a linear relationship between input features and the target variable.
- Support Vector Machine (SVM): A kernel-based model capable of capturing non-linear relationships in the data.

E. Model Training and Evaluation

The dataset was partitioned into a training set (80%) and a testing set (20%). Models were trained on the training partition and evaluated on the held-out test set using the following metrics:

- **R² Score:** Measures the proportion of variance in sleep quality explained by the model.
- **Mean Squared Error (MSE):** Measures the average squared deviation between predicted and actual values.

VI. PROPOSED SYSTEM

The proposed system is a machine learning-based framework designed to estimate sleep quality from individual smartphone usage patterns. It consists of three primary modules:

A. Data Acquisition Module

This module collects user input regarding smartphone usage through surveys or application interfaces, capturing daily screen time, nighttime usage duration, social media engagement frequency, notification counts, and sleep duration.

B. Data Analysis Module

This module performs preprocessing (missing value handling, normalization, encoding, and outlier removal), feature selection to isolate the most sleep-relevant variables, and applies machine learning algorithms—primarily Random Forest Regression—to identify usage patterns predictive of sleep quality.

C. Prediction Output Module

The final module generates a quantitative sleep quality score or categorical rating (Poor / Average / Good / Excellent) based on the processed input features. It also provides explanatory insights into how specific usage behaviors contribute to the predicted outcome, enabling users to make informed adjustments to their smartphone habits.

D. System Workflow

The end-to-end workflow proceeds as follows: (1) user input collection, (2) data preprocessing, (3) feature transformation and selection, (4) model training and inference, (5) sleep quality prediction, and (6) result presentation with actionable insights.

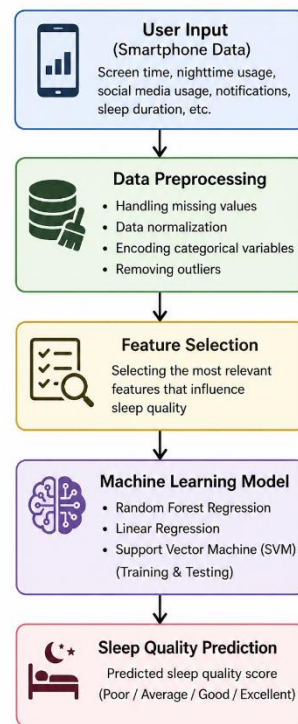


Fig. 1. End-to-end workflow of the proposed machine learning model

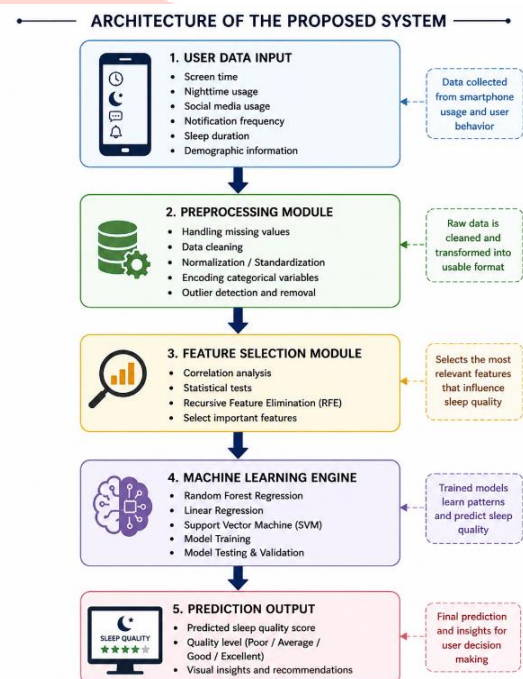


Fig. 2. Architecture of the proposed sleep quality prediction system

VII. EXPERIMENTAL SETUP

A. Dataset Configuration

The dataset encompasses self-reported smartphone usage and sleep data from participants aged 18–25. An 80/20 train-test split was applied to ensure that model performance was evaluated on unseen data.

B. Implementation Environment

All experiments were conducted in Python using the following libraries: Pandas for data manipulation, NumPy for numerical computation, and Scikit-learn for model

implementation. Development was performed using Jupyter Notebook and Google Colab.

C. Experimental Procedure

The experimental pipeline comprised the following steps:

- Data collection and preparation.
- Data cleaning and feature selection.
- Train-test dataset splitting.
- Model training using Random Forest Regression, Linear Regression, and SVM.
- Prediction generation on the test set.
- Performance evaluation using R^2 score and MSE.

VIII. RESULTS AND ANALYSIS

A. Model Performance Comparison

The predictive performance of all three models was assessed using the R^2 score. Results are presented in Table II.

TABLE II: MODEL PERFORMANCE COMPARISON BASED ON R^2 SCORE

Model	R^2 Score
Linear Regression	0.65
Support Vector Machine (SVM)	0.72
Random Forest Regression	0.82

The Random Forest Regression model achieved the highest R^2 score of 0.82, indicating strong explanatory power over sleep quality variance. Linear Regression performed poorly ($R^2 = 0.65$) due to its inability to model complex, non-linear behavioral interactions. SVM achieved a moderate performance ($R^2 = 0.72$).

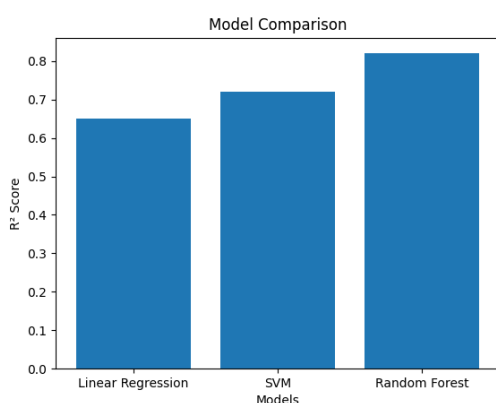


Fig.3. R^2 score comparison across all machine learning models.

B. Feature Importance Analysis

Feature importance scores derived from the Random Forest model are presented in Table III, highlighting the relative contribution of each variable to sleep quality prediction.

TABLE III: FEATURE IMPORTANCE ANALYSIS USING RANDOM FOREST MODEL

Feature	Importance Score
Screen Time	0.35
Nighttime Usage	0.30
Notification Frequency	0.20
Social Media Usage	0.10
Sleep Duration	0.05

Screen Time (0.35) and Nighttime Usage (0.30) were the most influential predictors, collectively accounting for 65% of model explanatory power. Notification Frequency (0.20) was the third most important feature, while Social Media Usage (0.10) and Sleep Duration (0.05) had comparatively lower impacts.

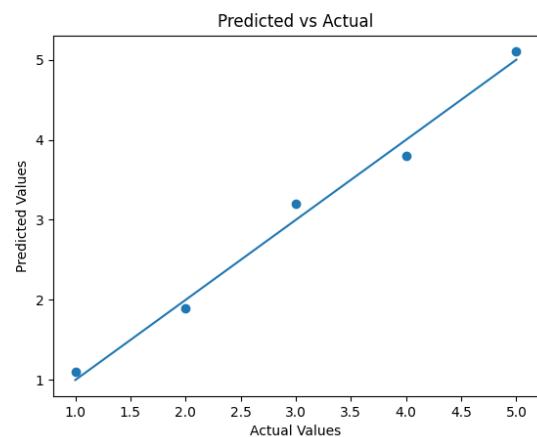


Fig.4. Predicted vs. actual sleep quality values (Random Forest Regression). The diagonal represents perfect prediction.

C. Error Analysis

Mean Squared Error (MSE) analysis confirmed that the Random Forest model produced the smallest prediction errors among all evaluated models, further establishing its superiority for this task. The ensemble nature of Random Forest allows it to handle multi-dimensional behavioral data more effectively than single-model approaches.

D. Key Findings

The study identified the following principal findings:

- Screen Time: Extended daily smartphone usage is strongly correlated with reduced sleep duration and degraded sleep quality.
- Nighttime Usage: Pre-sleep smartphone use delays sleep onset and disrupts sleep cycle progression.
- Notification Frequency: High notification rates contribute to sleep fragmentation and difficulty initiating sleep.
- Social Media Use: Frequent engagement with social media platforms increases pre-sleep cognitive arousal, impeding relaxation.

E. Discussion

The findings corroborate existing literature indicating that smartphone usage behaviors are significant determinants of sleep quality. The superior performance of the Random Forest model demonstrates the value of ensemble learning in capturing complex, non-linear relationships within behavioral datasets. The proposed framework offers a

reliable, scalable approach to sleep quality analysis grounded in data-driven methodology.

IX. OBSERVATIONS

The following key observations emerged from the experimental analysis:

- Data quality significantly impacts model accuracy; rigorous preprocessing is essential for reliable predictions.
- Clear behavioral patterns exist between smartphone usage metrics and sleep quality indicators.
- Nighttime smartphone usage and high notification frequency are the most disruptive behaviors for sleep health.
- Random Forest outperformed other models in capturing complex, multi-variable behavioral relationships.
- Targeted feature selection enhances predictive accuracy by focusing the model on the most sleep-relevant variables.
- Smartphone usage patterns and sleep quality are closely and consistently linked across the study population.

X. ADVANTAGES

- Provides reliable sleep quality predictions through machine learning-based behavioral analysis.
- Delivers actionable insights into the relationship between smartphone usage and sleep health.
- Supports digital wellness initiatives by promoting healthier smartphone usage habits.
- Can be integrated into web-based or mobile applications for real-time monitoring and personalized feedback.
- Ensemble modeling confers robustness against overfitting and noise in behavioral data.

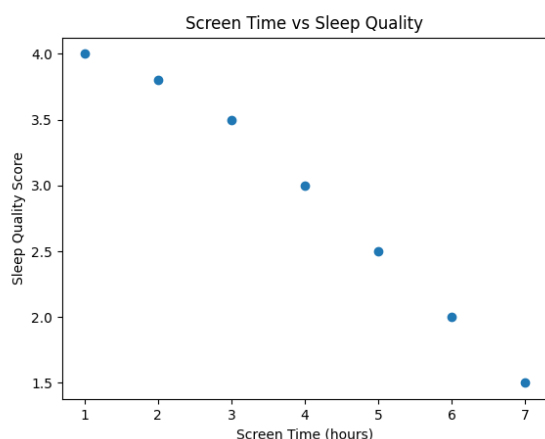


Fig. 5. Scatter plot of daily screen time (hours) vs. sleep quality score, illustrating the negative relationship between the two variables.

XI. LIMITATIONS

The proposed system has the following limitations:

- The dataset is relatively small, which may limit the model's generalizability to broader and more diverse populations.
- Reliance on self-reported data introduces potential for response bias and inaccuracy.
- The system does not support real-time data collection, restricting its application in continuous monitoring contexts.
- External factors such as psychological stress, physical environment, and lifestyle variables were not incorporated, which may also influence sleep quality.

XII. FUTURE SCOPE

The following enhancements are proposed for future work:

- Integration with wearable biosensors (e.g., fitness trackers, smartwatches) to enable objective, real-time data collection.
- Development of real-time sleep monitoring systems with continuous feedback capabilities.
- Application of deep learning architectures (e.g., LSTM, transformer-based models) for improved temporal pattern recognition.
- Development of personalized mobile health applications with adaptive recommendations.
- Incorporation of additional contextual variables such as physical activity levels, dietary habits, and psychosocial stress indicators.

XIII. CONCLUSION

This study presented a machine learning-based framework for predicting sleep quality from smartphone usage patterns among students and young adults. A dataset encompassing screen time, nighttime usage, social media engagement, notification frequency, and sleep duration was collected from participants aged 18–25. Three machine learning models were evaluated, with Random Forest Regression achieving the best performance ($R^2 = 0.82$), confirming its effectiveness in modeling complex behavioral relationships.

The findings demonstrate a strong association between excessive nighttime smartphone use and reduced sleep quality, underscoring the importance of responsible digital behavior for overall health and well-being. The proposed system provides a scalable, data-driven tool for sleep quality assessment and supports the broader objectives of digital wellness. Future advancements incorporating real-time monitoring, wearable sensor integration, and deep learning techniques hold the potential to further enhance the system's impact and applicability.

. REFERENCES

- [1] N. Cain and M. Gradisar, "Electronic media use and sleep in school-aged children and adolescents," *Sleep Medicine*, vol. 11, no. 8, pp. 735–742, 2010.
- [2] L. Hale and S. Guan, "Screen time and sleep among school-aged children and adolescents: A systematic literature review," *Sleep Medicine Reviews*, vol. 21, pp. 50–58, 2015.
- [3] A. M. Chang et al., "Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1232–1237, 2015.
- [4] J. C. Levenson, A. Shensa, J. E. Sidani, et al., "The association between social media use and sleep disturbance among young adults," *Preventive Medicine*, vol. 85, pp. 36–41, 2016.
- [5] L. Exelmans and J. Van den Bulck, "Bedtime mobile phone use and sleep in adults," *Social Science & Medicine*, vol. 148, pp. 93–101, 2016.
- [6] S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults," *BMC Public Health*, vol. 11, no. 66, pp. 1–11, 2011.
- [7] R. Lemola, M. Perkinson-Gloor, S. Brand, J. F. Dewald-Kaufmann, and E. Grob, "Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age," *Journal of Youth and Adolescence*, vol. 44, no. 2, pp. 405–418, 2015.
- [8] H. Lund, B. Reider, A. Whiting, and J. Prichard, "Sleep patterns and predictors of disturbed sleep in a large population of college students," *Journal of Adolescent Health*, vol. 46, no. 2, pp. 124–132, 2010.
- [9] M. A. Gradisar, G. Wolfson, L. Harvey, et al., "The sleep and technology use of Americans: Findings from the National Sleep Foundation," *Sleep Health*, vol. 1, no. 2, pp. 143–147, 2013.
- [10] A. Punamäki, L. Wallenius, K. Nygård, H. Saarni, and H. Rimpelä, "Use of information and communication technology and perceived health in adolescents," *Scandinavian Journal of Work, Environment & Health*, vol. 33, no. 2, pp. 87–99, 2007.
- [11] J. A. Exelmans and J. Van den Bulck, "The use of media as a sleep aid in adults," *Behavioral Sleep Medicine*, vol. 14, no. 2, pp. 121–133, 2016.
- [12] M. Van den Bulck, "Television viewing, computer game playing, and Internet use and self-reported time to bed and time out of bed in secondary-school children," *Sleep*, vol. 27, no. 1, pp. 101–104, 2004.
- [13] S. Higuchi, Y. Motohashi, Y. Liu, and A. Maeda, "Effects of playing a computer game using a bright display on presleep physiological variables, sleep latency, slow wave sleep and REM sleep," *Journal of Sleep Research*, vol. 14, no. 3, pp. 267–273, 2005.
- [14] C. Cajochen, "Alerting effects of light," *Sleep Medicine Reviews*, vol. 11, no. 6, pp. 453–464, 2007.
- [15] A. K. Wheaton, A. M. Ferro, and J. B. Croft, "School start times, sleep, behavioral, health, and academic outcomes: A review of the literature," *Journal of School Health*, vol. 85, no. 5, pp. 363–381, 2015.